Chapter 5: Least Squares Regression
Key Ideas

- The **least-squares regression line** of y on x is the line that makes the sum of the squares of the **vertical distances** of the data points from the line as small as possible.
- **The least-squares regression line equation:**

$$\hat{y} = a + bx$$

with

slope $b = r\dfrac{s_y}{s_x}$ and intercept $a = \bar{y} - b\bar{x}$.

- The square of the correlation, r^2, is the **fraction of the variation** in the values of y that is **explained** by the least-squares regression of y on x.
- **Note:** Computations may be carried out using technology (e.g., SPSS)

- A **residual** is the difference between an observed value of the response variable and the value predicted by the least-squares regression line.

$$\text{residual=observed-expected}$$
$$= y - \hat{y}$$

- A **residual plot** of the residuals against the x values helps assess the fit of a regression line.

- Scatter plots (y vs x) and residual plots may uncover noteworthy observations:

  - **Outlier**: an observation that lies outside the overall pattern of the other observations.

    - Outliers have unusually large positive or negative residuals.

  - **Influential observation**: an observation whose removal would markedly change the equation of the regression line.

    - Points that are extreme in the x direction are often influential.